



Centre de Recherche en Economie et Droit de l'Energie

**EXPLAINING THE CONVENIENCE YIELD IN THE WTI CRUDE OIL MARKET  
USING REALIZED VOLATILITY AND JUMPS**

**Benoît SÉVI**

**Cahier de recherche n° 14.10.109**

1 octobre 2014

**Centre de Recherche en Economie et Droit de l'Energie**

CREDEN – Equipe ART Dev – Université Montpellier 1  
Faculté d'Economie, Av. Raymond Dugrand, C.S. 79606  
34960 Montpellier Cedex 2, France  
Tél. : +33 (0)4 34 43 25 04

# Explaining the convenience yield in the WTI crude oil market using realized volatility and jumps\*

**Benoît Sévi**<sup>†</sup>

Grenoble Alpes Université, UMR 1215 GAEL, 38000 Grenoble, France

October 1, 2014

---

\*This research was initiated while the author was a visiting scholar at the London Business School (MSO Department), whose support and hospitality have been greatly appreciated. Comments by Derek Bunn are gratefully acknowledged as well as very detailed comments by an anonymous referee which help to improve significantly the paper.

<sup>†</sup>Email: benoit.sevi@gmail.com

## **Abstract**

In this paper, we first provide empirical evidence of the existence of intraday jumps in the crude oil price series. We then show that these jumps, in conjunction with realized volatility measures, are important in modeling the convenience yield over the 2001-2010 period. Our empirical results indicate that lagged jump mean only explains around 16% of the weekly convenience yield. Our best specification, including variation in inventories, eight-week realized variance and the 250-day jump mean is able to explain around 61% of the weekly convenience yield. Importantly, our results are not driven by the simultaneous determination of the various variables at work as we only use lagged variables in all regressions.

*JEL Classification:* C15, C32, C53, G1, Q4.

*Keywords:* convenience yield, realized volatility, jump, inventory.

# 1 Introduction

Our paper aims at modelling the convenience yield using measures of volatility and jumps computed using intraday data in the WTI oil futures market. We add these new measures to other measures previously used in the literature (see Pindyck (2004)) such as the 5-week moving average of volatility computed using daily returns, as well as inventory level and shocks to the spot price. We provide evidence that intraday data are helpful in modeling the 1-month convenience yield and that the distinction between volatility and jumps further increases the explanatory power.

Recent mathematical finance literature suggest to model the convenience yield as a mean-reverting process (see Gibson and Schwartz (1990), Schwartz (1997), Hilliard and Reis (1998), Schwartz and Smith (2000) and Casassus and Collin-Dufresne (2005) among the most representative contributions). Liu and Tang (2011) propose an affine model for the oil price with three state variables (spot price, interest rate, and the convenience yield). Their model is able, among other things, to capture an essential characteristic of oil futures, i.e. the positive relationship between volatility and the convenience yield. Mirantes et al. (2013) develop a four-factor model and emphasize the seasonal feature of the convenience yield in addition of its stochastic behavior.

The stochastic nature of the convenience yield calls attention to its modelling, what we attempt in the present paper. Gorton et al. (2013) show that the convenience yield is a decreasing and non-linear function of inventories. As such, inventories may have predictive power for the convenience yield. Also, Pindyck (1994, 2001, 2004) shows that the convenience yield should be a function of the level of spot price, the level of inventories and the level of volatility.<sup>1</sup> Because all these variables are jointly determined, we only use in our empirical applications lagged explanatory variable to avoid the simultaneity issue. Other papers modeling the convenience yield using similar variables are Chiou Wei and Zhu (2006) and Borak et al. (2006), but none use realized measures and their explanatory variable for the volatility component is generally not significant.

This paper is the first to model the convenience yield using the rich information in high-frequency data. Existing work uses at best daily data. Because realized estimators provide a more robust measure of the latent volatility, we expect to improve empirical results by using these estimates. As such, we extend the analysis in Pindyck (1994, 2004).<sup>2</sup> In addition, we distinguish between the

---

<sup>1</sup>The general relationship between inventories, volatility, prices and the convenience yield is nicely presented in Geman (2005), but using informal arguments. Knittel and Pindyck (2013) and earlier references cited therein provide theoretical support for such a relationship.

<sup>2</sup>In contrast with Pindyck (1994), Pindyck (2004) takes explicitly account for price volatility as a factor determining the convenience yield.

information in the continuous component and the jump component from high-frequency prices. This distinction has proven to be very useful in the volatility forecasting literature (see Andersen et al. (2007)) and we investigate its usefulness here for modeling the convenience yield.

We consider the weekly 1-month convenience yield for the most traded commodity futures contract in the world, namely the WTI crude oil futures from CME-NYMEX. Our main results are as follows. First, realized volatility measures computed over the last  $i$  weeks for  $i = 1, \dots, 8$  have a good explanatory power, better than the 5-week measure using daily data in Pindyck (2004) when  $i$  equals 5 or above. Second, the explanatory power of the jump mean statistic alone is around 16%. As jumps are representative of large and rapid changes in oil futures prices, our results point to a relationship between large intraday returns (jumps) and the convenience yield. As the latter describes needs and expectations of economic agents in the oil market, our study partly explains how investors react to large changes in oil futures prices through their trading strategy.

Our finding about the existence of a relationship between the 1-month convenience yield and lagged volatility and jumps has implications for the modeling of oil returns. Indeed, two- or three-factor models now systematically include the convenience yield risk as a risk factor to fit futures price data better.<sup>3</sup> As we show that the convenience yield risk is related to other risk factors such as jump and volatility risk, these risks should not be modeled independently from the convenience yield risk any more as is often the case in stochastic modeling of energy prices.

As a further implication, it should be emphasized that if a stochastic convenience yield leads to market incompleteness as pointed out in Geman (2005) then the link between volatility (or jumps) and the convenience yield may help to implement optimal hedging strategies using existing instruments such as the oil volatility index (OVX) that has been developed to track the model-free implied volatility for the WTI oil market. These hedging strategies would be developed on the basis of a statistical relationship between volatility and the convenience yield and would help to reduce the overall price exposure when no financial instrument exists to specifically deal with the convenience yield risk. This kind of strategy, coined as “cross-hedging”, relates to the early contribution of Anderson and Danthine (1981), where the authors show how an optimal hedging policy can be implemented using hedging tools whose underlying asset is not the same as the asset for which the original risk is bear.<sup>4</sup>

---

<sup>3</sup>Convenience yield risk refers to the risk of change in the convenience yield level, not the risk related to the volatility of the convenience yield which is a very different concept.

<sup>4</sup>The idea here would be to hedge the convenience yield risk using oil volatility futures (futures on OVX) exploiting the existing correlation between the convenience yield and the oil volatility that we highlight in the present study.

Our results may also be of interest for policy-makers as we emphasize a link between the volatility (and jumps) and the future convenience yield. Recall that the convenience yield is determined on the basis of subsequent futures prices. Our results thus show that volatility and jumps play a major role in explaining the term structure on the oil market. Moreover, the volatility has explanatory power for the future storage policy by firms as this storage policy explains the price gap between futures contracts of different maturities. Policy-makers may then influence the storage policy of firms in periods of high volatility, where the convenience is expected to increase, by managing oil strategic reserves according to their objectives.

## 1.1 Relevant literature

Following the recent availability of intraday data for most existing financial markets, and in particular commodity markets, recent research has investigated various properties of oil financial markets using this data. Wang et al. (2008) investigate the distributional properties of realized volatility and standardized returns in the WTI futures market over the 1995-1999 period. Tseng et al. (2009) use pseudo long memory time series models to evaluate the contribution of jumps to forecasting the conditional volatility in the WTI futures market for the 2000-2007 period. More recently, Liu and Wan (2012) study the long-range dependence in the Shanghai fuel oil futures market over the 2004-2011 period and emphasize the role of tick-by-tick data to forecasting conditional volatility. The authors also highlight the role of trader activity which has a very significant impact on the level of volatility in the futures market. Chevallier and Sévi (2012) empirically study the volatility-volume relationship using intraday data and show the role of jumps in bearing this relation over the period 2006-2010. Sévi (2014) provides an extensive study of the predictive properties of various HAR-class models for long 1987-2010 period and shows that the basic HAR model provides very good forecast performance out-of-sample relatively to more sophisticated models, thereby questioning the interest to disentangle the jump component from the continuous component for the purpose of forecasting oil price volatility. Finally, Baum and Zerilli (2013) use non-parametric estimates of jumps in the WTI oil markets to estimate the parameters of a continuous-time stochastic volatility model with jumps. The authors provide strong empirical evidence of the importance of jumps in adequately modeling the oil return process.

Other papers have investigated the issue of jumps using daily data. Those papers are Askari and Krichene (2008), Lee et al. (2010) and Gronwald (2012). Askari and Krichene (2008) estimate the

various components of a stochastic process and provide evidence of jumps. Lee et al. (2010) and Gronwald (2012) investigate and estimate the presence of jumps relying on the methodology in Chan and Maheu (2002). Our paper improves upon these last two by using nonparametric methods along with intraday data.

Our measures of jump distribution use the recent contribution by Tauchen and Zhou (2010) who suggest to use the characteristics of the distribution of jumps to predict credit spreads. To our best knowledge, these tools have only been used in Wright and Zhou (2009), Zhang et al. (2009) and Evans (2011) to date. The idea behind these papers is that characteristics of jump distribution may be helpful in explaining some financial variables such as credit spread (at the aggregate or firm level) or the bond premia because the returns are function of agents' aversion to large losses. We use the same measures as in Tauchen and Zhou (2010) along with measures of realized volatility to model the 1-month convenience yield.

Symeonidis et al. (2012) provide empirical evidence of the relationship between inventory and the shape of the forward curve. The convenience yield is related to the forward curve and, as such, the authors also analyze the link between inventory and the convenience yield. The authors also highlight the relationship between inventory and conditional volatility. In our investigation of the predictive properties of various predictors of the convenience yield, we also explore the forecasting performance of inventory and volatility. However, we distinguish between the continuous component and the jump component to enrich our set of predictors.

The convenience yield is also included in dynamic models of the oil storage as in Pieroni and Ricciarelli (2008) who extend the Pindyck (1994, 2001) model. Because all the variables such as the price, the volatility, the convenience yield and the inventory level are endogenously determined, we use lagged variables to deal with the issue of simultaneity.

In the next Section, we provide some details about the econometric approach. Section 3 presents the main empirical findings of our regression analysis and discusses some robustness checks that we also computed. Section 4 provides some concluding remarks and suggestions for future research about the empirical properties of the convenience yield.

## 2 Empirical approach

We present in this Section how we estimate realized quantities that will be used in the next Section for regression analysis. We first develop the concept of bi-power variation which is used along with realized volatility to detect jumps. Next, we explain the computation of jumps statistics.

### 2.1 Detecting jumps in oil returns

Jumps are defined as large returns that could be of interest to investors that face the convenience yield risk in commodity markets. Academic research in financial economics suggests that jumps generally are related to the resolution of uncertainty and should be greater the larger is the dispersion of beliefs about a given financial asset (see Giot et al. (2010) or Evans (2011) for a discussion). The recent availability of very high-frequency data allows to detect jumps statistically. Consider intraday returns  $r_{t,\Delta}$  that are the log-difference of the observed oil price with a sampling period of  $\Delta$  which could be one or a few minutes depending on the liquidity features of the market under consideration (see below). The day- $t$  daily realized variance (RV) (see Andersen and Bollerslev (1998)) is the sum of these squared intraday returns over the all day:

$$RV_{t+1}(\Delta) = \sum_{j=1}^{1/\Delta} r_{t+j,\Delta,\Delta}^2 \quad (1)$$

with, for simplicity,  $1/\Delta$  denotes an integer.

Consider jumps as dramatically large intraday returns. The realized variance is a measure of conditional variance that is not robust to jumps in the sense that jumps in intraday returns enter the formulae as all other returns do. Conversely, a measure of realized variance that is robust to jumps is the bi-power variation (BPV) defined in Barndorff-Nielsen and Shephard (2004) as the sum of products of pairs of consecutive intraday returns over day  $t$ :

$$BPV_{t+1}(\Delta) = \mu_1^{-2} \sum_{j=2}^{1/\Delta} |r_{t+j,\Delta,\Delta}| |r_{t+(j-1),\Delta,\Delta}| \quad (2)$$

with the scaling factor  $\mu_1 \equiv \sqrt{(2/\pi)} = E(Z)$  the mean of the absolute value of a standard, normally distributed random variable  $Z$ . The intuition behind this estimator is that very large returns (jumps) have asymptotically no impact on the realized measure because they are multiplied by



very small returns as the observation frequency increases.<sup>5</sup>

Barndorff-Nielsen and Shephard (2004) suggest to use a 'staggered version' of the BPV, that is to use products of intraday returns that are not exactly consecutive but separated by one or more returns, to deal with the well-known issue of serial correlation due to microstructure effects in financial markets. In what follows, we always use the one-period staggered version of the BPV estimator.

Barndorff-Nielsen and Shephard (2006) use the BPV estimator to test for the presence of jumps in intraday returns using the ratio test statistic defined as:

$$Z_{t+1}(\Delta) = \frac{\left(1 - \frac{BPV_t(\Delta)}{RV_t(\Delta)}\right)}{\sqrt{(\mu_1^{-4} + 2\mu_1^{-2} - 5)\frac{1}{1/\Delta} \max(1, TQ_t(\Delta)/BPV_t^2(\Delta))}} \quad (3)$$

where  $TQ_{t+1}(\Delta)$  is the tri-power quarticity which is a robust-to-jumps estimate of the integrated quarticity and is computed as follows:

$$TQ_{t+1}(\Delta) = \Delta^{-1} \mu_{4/3}^{-3} \sum_{j=3}^{1/\Delta} |r_{t+j,\Delta,\Delta}|^{4/3} |r_{t+(j-1),\Delta,\Delta}|^{4/3} |r_{t+(j-2),\Delta,\Delta}|^{4/3} \quad (4)$$

where  $\mu_{4/3} = 2^{2/3} \cdot \Gamma(7/6) \cdot \Gamma(1/2)^{-1} = E(|Z|^{4/3})$ . The ratio test statistic is asymptotically normally distributed and has very good small-sample properties with respect to most widely used stochastic processes as shown in Huang and Tauchen (2005) in a large-scale Monte-Carlo experiment.

When this statistic is significant, then the difference between the RV and the BPV is too large to be a simple realization of a Brownian motion and should be considered as a jump. Thus, the jump component is given by:

$$J_{t+1,\alpha}^{BPV}(\Delta) = [RV_{t+1}(\Delta) - BPV_{t+1}(\Delta)] \times I[Z_{t+1}(\Delta) > \Phi_\alpha] \quad (5)$$

where  $I[\cdot]$  is the indicator function which identifies the significance of the  $Z_{t+1}(\Delta)$  statistic in excess of some critical value of the Gaussian distribution  $\Phi_\alpha$ .

<sup>5</sup>Other measures of realized variance that are robust to jumps have been suggested in the literature. Some of them are used for robustness analysis in the penultimate section. A nice presentation of these measures as well as their small sample properties for the sake of detecting jumps are discussed at length in Theodosiou and Žikeš (2011) and Dumitru and Urga (2012).

## 2.2 Realized jumps statistics

Tauchen and Zhou (2010) suggest to use the characteristics of the distribution of realized jumps over a given time period. Using a sampling interval  $\Delta$  for intraday returns, a significance threshold  $\alpha$  and the BPV estimator, realized jump for day  $t$  is estimated as:

$$RJ_{t+1,\alpha}^{BPV}(\Delta) = \text{sign}(r_t) \times \sqrt{J_{t+1,\alpha}^{BPV}(\Delta)}$$

with  $r_t$  the day  $t$  daily return which, following Tauchen and Zhou (2010) gives its sign to the realized jump. The idea behind this formulae is that an important jump should be, most of the time, of the same sign as the daily return.

The jump intensity, the jump mean and the jump standard deviation are defined as follows:

$$\hat{\lambda}_J = \frac{\text{number of realized jumps}}{\text{number of total trading days}}$$

$$\hat{\mu}_J = \text{mean of realized jumps}$$

$$\hat{\sigma}_J = \text{standard deviation of realized jumps}$$

These characteristics, when computed using S&P 500 returns, are shown to help explaining and forecasting AAA and BAA bond spreads better than interest rate volatility which is known as a good predictor in the existing literature (see Tauchen and Zhou (2010)). Moreover, the authors provide strong evidence of the robustness of their statistics by means of extensive simulations.

In what follows, these statistics will be used to model the convenience yield in the WTI crude oil futures market as we believe that exceptionally large returns may convey some information about the level of the convenience yield.

## 3 Empirical findings

### 3.1 Data

We use intraday data for the WTI New York Light Sweet Crude Oil futures contract which is traded on the New York Mercantile Exchange (NYMEX), now part of the CME Group. Our data include all transactions of the open outcry session over the period going from October 5, 2001 to January 15, 2010 resulting in 2071 trading days after removing days with (1) transactions outside the official trading period, (2) transactions with a variation of more than 5% in absolute value compared to the previous transaction and (3) transactions not reported in chronological order. We also remove days with insufficient trading activity. In particular, we do not further consider days with a shortened trading period and days with more than ten zero-return. We end with trading days where liquidity is sufficient to ensure that our estimates of realized variance and jumps are consistently estimated. Our weekly series consists in 382 weekly observations where daily realized variance are aggregated at the weekly frequency.

We build continuous time series by using front-month contracts and the next-to-delivery futures contract, and by switching from one contract to the next as soon as the volume for the next month is higher than the volume for the present contract.<sup>6</sup> These futures were selected primarily because they are the most actively traded futures in their own category, so the problem of infrequent trading is minimized.

We present a plot of the nominal oil price in Figure 1 along with a plot of the returns in Figure 2. We observe the well-known boom-and-bust of the late 2008 along with quite large nominal prices in the last months of the sample. Moreover, from the return Figure we note that the volatility is much larger during the drop period. Volatility is also strongly time-varying with clusters of large (absolute) returns in some specific periods indicating significant heteroscedasticity.

#### 3.1.1 Optimal sampling frequency and jump detection threshold

We compute realized variance using the widely-used 5-minute sampling interval which is a good trade-off between removing most of the effects of a possible microstructure noise and keeping

---

<sup>6</sup>We checked that in all cases, the contract with the highest volume also is the contract with the highest number of transactions thereby responding to the possibly different logics of the financial literature on the information flow that could be proxied either using volume or the number of trades.

in hand most of the available observations. Following Chevallier and Sévi (2012), we computed some volatility signature plots (i.e. plots of the realized variance for a set of sampling intervals) for the extended 2001-2010 period and show the stability of the realized variance estimator even at very high frequencies.<sup>7</sup> As such, our 5-minute choice appears to be a conservative choice with regards to the discarded data but remains in line with existing literature and should allow easier comparison with existing studies.<sup>8</sup>

Realized volatility (the square root of the realized variance) in annualized terms is plotted in Figure 3. The Figure confirms the heteroscedastic behavior of oil returns. The average volatility level is around 50% which is much higher than for standard financial assets. Interestingly, days with large drops can lead to very different realized volatility estimates as exemplified in the start of 2003 and the end of 2008, depending on the intraday trading profile of the day of interest. Jumps that are detected at the 0.1% level are plotted in Figure 4. This statistical threshold for jump detection is common in financial econometric literature (see Andersen et al. (2007) among many others) as it permits to only detect very significant jumps that generally are of primary interest for empirical applications. Less significant jumps may have an impact on the characteristics of the realized jump distribution and make the results from the regression analysis less clear-cut. Surprisingly, we notice from Figure 4 that realized jumps are not often associated with turbulent periods. In particular, realized jumps may be absent in times of large realized volatility and *vice versa*. This indicates that the jump and the continuous component are empirically found to be very different in their dynamics and may well have a different information content.

At the 0.1% detection level, we obtain a contribution of realized jumps to the total realized volatility of 8.27% which is larger than for existing studies. For instance, for the S&P 500 futures, the contribution of realized jumps is around 5%.<sup>9</sup>

### 3.1.2 Computing the convenience yield

As the convenience yield is our endogenous variable in all regressions, it is worth explaining how it is computed as various definitions may be used at this step. Let  $S$  be the spot price at day  $t$ .

<sup>7</sup>Volatility signature plots are available upon request to the authors.

<sup>8</sup>Liu et al. (2012) provide an extensive study of the forecasting performance of linear time series models using estimates of realized volatility that are computed at various frequencies. They conclude that the 5-minute sampling interval is generally difficult to beat when forecasting the conditional volatility.

<sup>9</sup>The contribution of jumps to the total realized variance is 9.55% and 10.72% for the 1% and the 5% level respectively. Hence, we observe a relative stability of the contribution of jumps for the various thresholds. This leads to consider that empirical results might be relatively similar for these different thresholds.

For storable commodities, the futures price is given by the well-known cash-and-carry arbitrage relationship :

$$F_{S,T} = S e^{(r-cy)(T-t)} \quad (6)$$

with  $F_{S,T}$  the futures price with maturity  $T$ ,  $r$  the risk-free rate and  $cy$  the convenience yield for the  $[t, T]$  period. We follow Gibson and Schwartz (1990) and compute each week the annualized forward convenience yield as:

$$cy_{T-1,T} = r_{T-1,T} - 12 \ln \left[ \frac{F_{S,T}}{F_{S,T-1}} \right] \quad (7)$$

with  $F_{S,T-1}$  and  $F_{S,T}$ , the nearest-to-maturity and second nearest-to-maturity futures contracts respectively. The front-month futures price is taken as a proxy for the spot price as the latter is often influenced by delivery issues.  $r_{T-1,T}$  is the 1-month risk-free rate proxied by the T-bill rate that is available from the Federal Reserve Board statistics release database. We only consider the 1-month convenience yield, *i.e.* that the time between  $T$  and  $T - 1$  is 1 month. Other durations could be examined to gain further insights about the term structure but this is a quite different exercise which is left for future work.

The resulting convenience yield is represented in Figure 5. We observe a very large negative spike in the convenience yield during the 2008-bust period where futures prices were large with respect to the spot price as participants in the market were still expecting prices to recover higher levels.

Descriptive statistics for the convenience yield that is a stationary variable are provided in the first column of Table 1.

### 3.2 Modelling the convenience yield: main findings

We model the weekly convenience yield in level using variables that have been used in previous studies along with variables that are derived from intraday data as a new contribution to the literature. In addition to change in inventory data, we use conditional volatility and jump statistics to model the convenience yield.

To keep things relatively conditional, we use the 8-week realized variance but we do not extend

the analysis to longer horizons. Next, we compute jumps over the last year of trading (250 days) when Tauchen and Zhou (2011) use a two-year period.<sup>10</sup>

As shortly noticed in the Introduction, we only use lagged explanatory variables to deal with the simultaneity issue. Indeed, it is natural to think that variables such as volatility, inventories, prices and the convenience yield are jointly determined. Models such as in Pindyck (1994, 2001, 2004), Knittel and Pindyck (2013) are in line with this idea that is well-known in the academic literature about commodities (see Geman (2005) for an illustration). As such, as in Zhang et al. (2009), we avoid to use variables that are contemporaneous with the convenience yield as the explanatory power in our regressions may be significantly inflated due to the joint determination of these variables.

Also, it should be noted that the spot price shock used in Chiou Wei and Zhu (2002) and Pindyck (2001) is not significant in all our regressions and thus the results are not reported. This points to the importance of the time period under consideration as in past studies the spot price shock was significant.

### 3.2.1 Univariate regressions

We first investigate the explanatory power of each individual exogenous variable. Estimates are provided in Table 2. Estimation is by OLS.<sup>11</sup> We observe that the variable that has the best explanatory power is the change in inventory level (regression 1) with a high  $R^2$  of 35.47%.<sup>12</sup> The variable is highly significant at the 1% level and the estimate has the expected sign, which is negative. From this regression, we conclude that the convenience yield and changes in the inventory level are negatively correlated as is implied from the theoretical literature.

Regressions 2 and 3 provide evidence of the significance of the two explanatory variables representing the conditional volatility.<sup>13</sup> However, despite both are highly significant, the 8-week volatility based on the sum of daily realized variances has a better explanatory power. The im-

<sup>10</sup>We also run regressions using jumps statistics computed over different time periods (60 or 120 days) with no success thereby emphasizing the need of a sufficient period of time to obtain robust information about jumps characteristics as coined in Tauchen and Zhou (2011). Recall that jumps characteristics are descriptive statistics in essence and are estimated with better precision when the time period grows.

<sup>11</sup>We replicated the all analysis (univariate and multivariate regressions) using GMM to deal with the endogeneity and “error-in-variables” issues following Pindyck (2004). Our results are similar in all cases and available upon request to the authors.

<sup>12</sup>Inventory level is plotted in Figure 7 and taken in first difference.

<sup>13</sup>Because volatility predictors as well as jump predictors are highly persistent (see Table 1) we use as the regressor the residual from an AR(12) which is found to be sufficient to remove serial correlation for all series.

provement coming from using intraday data is not so high but results in more than a 6% raw increase of the  $R^2$  corresponding to an improvement of the explanatory power of almost 25%. Overall, conditional volatility is, with the change in inventory level, a good predictor for the future convenience yield. The importance of intraday data may also be considered with respect to the contribution by Chiou Wei and Zhu (2006) who model the convenience yield for the U.S. natural gas futures market and do not find significant estimate for their measure of conditional volatility. It might be that using daily data is not enough to recover a robust relationship between the convenience yield and the conditional volatility in this market.

We now turn to the explanatory power of variables that are based on jump characteristics. From regression 4, we note that the jump mean is significant at the 1% level and should be considered as a good explanatory variable for the modeling of the convenience yield. The estimated value is positive indicating that an increase in the jump mean has a positive impact on the convenience yield, similar to the impact of volatility we discussed above. Recall that we detect the jumps using the methodology in Tauchen and Zhou (2011) that allows to sign each jump with reference to the daily return. As such, a negative jump mean then reduces the convenience yield. This is expected from theory where decreases in price are associated with lower convenience yield. Importantly, the estimate is highly significant which indicates a possibly important role for large changes in oil prices beyond the information that is already present in standard measures of conditional volatility. The estimate for the jump volatility also is highly significant indicating that an important variability in large (positive or negative) returns has an impact on the 1-month convenience yield. The core interest to consider jump characteristics in modeling the convenience yield is now investigated by means of multivariate regressions.

### **3.2.2 Multivariate regressions**

Table 3 reports results from multivariate regressions to gauge the relative importance of our set of regressors. Regression 1 shows that the information in the jump mean variable seems to subsume the information content in the jump volatility variable, which becomes insignificant when the jump mean variable is present. The jump volatility also is insignificant when used along with the inventory variable, as shown in regression 3.

An important result is in regression 5 where the measure based on daily data over the last 5 weeks following Pindyck (2004) is compared with our measure of volatility derived from intraday data

over the last 8 weeks. We observe that the measure of conditional volatility using high-frequency data provides better explanatory power and makes insignificant the genuine measure based on daily closing prices. This result was expected both from the theoretical properties of the realized variance estimator and from the numerous studies which show the superiority of intraday data in several economic applications (Andersen and Benzoni (2009) present most of them).

Regression 4 provides excellent results with a  $R^2$  as high as 57.73% using inventory and volatility from intraday data as regressors. The significance of the conditional volatility in this regression shows that changes in inventories does not subsume the information content of the conditional volatility. This is an important result as these two variables are jointly determined and might not convey significantly different information. We observe that this is not the case and that the information of each variable plays a role in explaining the convenience yield. Our preferred specification is regression 6 where we obtain the highest adjusted  $R^2$  with a value of 61.31%. In this regression, changes in inventory, realized volatility over the last 8 weeks and the jump mean are all statistically significant. This is an interesting result in light of the results in regression 2 where the jump mean is not more significant when used along with the inventory variable. It appears that adding the conditional volatility helps in recovering the significance of the jump mean.

Overall, our results point to the importance of using intraday data to obtain better estimates of conditional volatility that have higher explanatory power for the convenience yield. Also, we show the additional information content of jump mean beyond what are able to explain traditional variables such inventories or conditional volatilities.

### 3.3 Sensitivity analysis

We suggest to analyze the robustness of our empirical results through three dimensions. First, we modify the threshold for the jump detection test in Eq. 3. Second, we use an alternative estimator of realized quantities (realized variance, etc.) to better mitigate the impact of jumps and zero-returns. Third, we investigate the relevance of various horizons to estimate the conditional volatility.



### **3.3.1 Jump detection threshold**

We check the robustness of our regression analysis using alternative thresholds for the detection of jumps. In particular, we check whether identifying jumps using a lower threshold leads to comparable results. Jumps detected at the 5% level are plotted in Figure 6. We observe obvious similarities between these realized jumps and those that are obtained using the 0.1% level. As noted earlier, the contribution of jumps to the realized volatility is quite similar for both significance thresholds. Unreported results that are available upon request provide evidence of the same relationship between the convenience yield and realized jump mean and volatility in the case of the 5% level.

### **3.3.2 Microstructure noise robust estimation of realized quantities**

In addition to the staggered version of the BPV estimator that is used in this paper following empirical arguments in Huang and Tauchen (2005) to deal with the microstructure noise issue, we use the two-scale estimator developed in Zhang et al. (2005) to compute both the RV and the BPV. As a further robustness check, we also use the median realized volatility (MedRV) estimator developed in Andersen et al. (2012). This estimator has much better properties than the BPV (see Dumitru and Urga (2012)) for detecting jumps and behaves much better in settings where the number of zero-return is significant and where jumps are present in the data. Results are similar to those using standard estimators and not reported to save space.

### **3.3.3 Selected horizon for the computation of conditional volatility measures**

As a final sensitivity analysis, we investigate the importance of the time period used to compute measures of volatility. In Table 4, we report results from univariate regressions of the 1-month convenience yield using measures of realized volatility that are computed using horizons going from 1 to 8 weeks. We observe that all reported estimates are significant, often at the 1% level.

We investigated various horizon lengths for the use of daily data and the best results are for the 5-week horizon which is also used in Pindyck (2004). We recall the result for this estimate of conditional volatility in Table 4. It seems that we have a gain in enlarging the set of observations. In particular, the best results are found for 6 to 8 weeks of high-frequency data. The best explanatory power is for the 8-week horizon to be compared with the best result for daily data at the 5-week

horizon. The explanatory power then decreases for larger horizons.

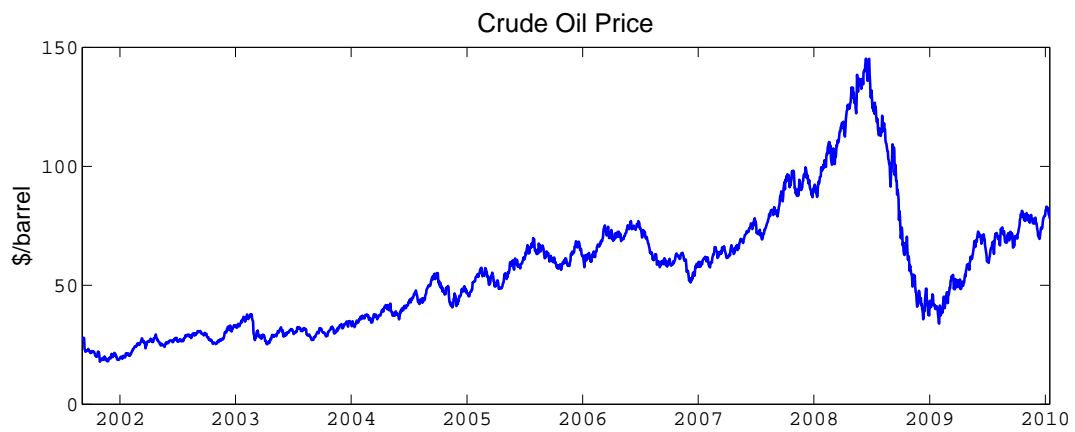
Overall our robustness checks validate our empirical findings and confirm the established relationship between the convenience yield and the various explanatory variables.

## 4 Conclusion

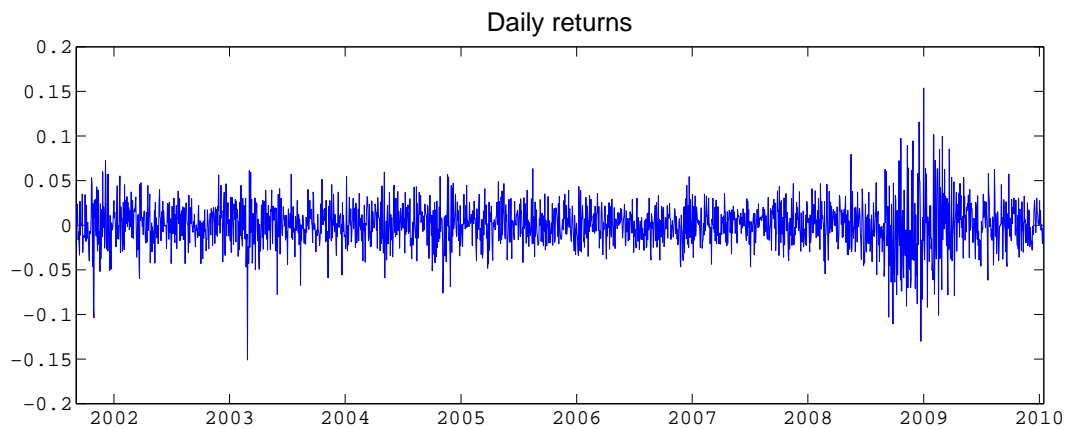
This paper confirms that the convenience yield dynamics can be modeled using variables such as inventories, conditional volatility and jumps. Moreover, it is shown that the information content in intraday data significantly helps in modeling the weekly 1-month convenience yield over the 2001-2010 period. As the convenience yield is a major input in stochastic processes that are used for modeling oil returns, our empirical result has important implications in this respect. In particular, it appears that volatility, jumps and the convenience yield are not independent quantities and risk factors associated with these quantities should exhibit this dependence.

A recent literature has emerged showing that oil prices help to predict financial returns (Pollet (2003), Driesprong et al. (2008)). Refining this approach, Gospodinov and Ng (2013) provide empirical evidence that the convenience yield in commodity markets is able to predict not only commodity returns but also financial asset returns. From this result, a natural question arises that may be of interest for future work: have realized volatility and jump distribution statistics any explanatory power for future oil returns and more generally for stock and bond returns?

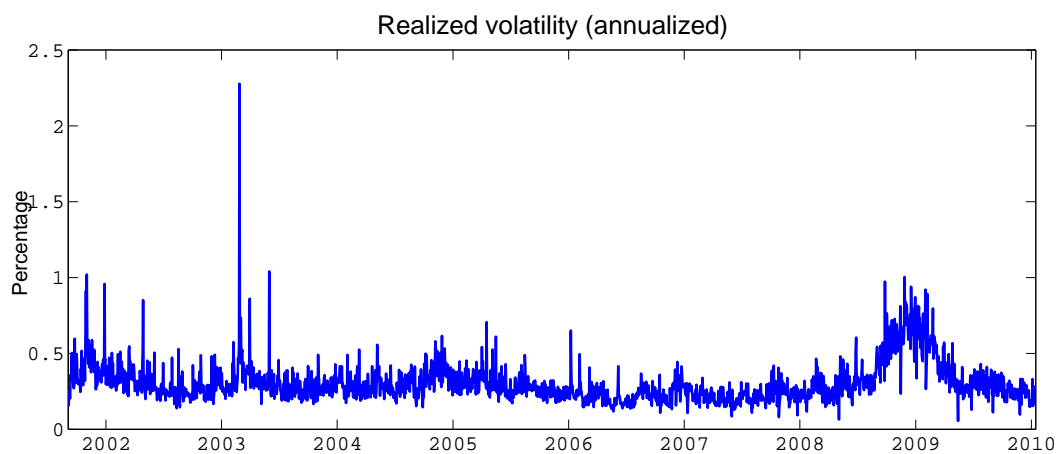
## Figures



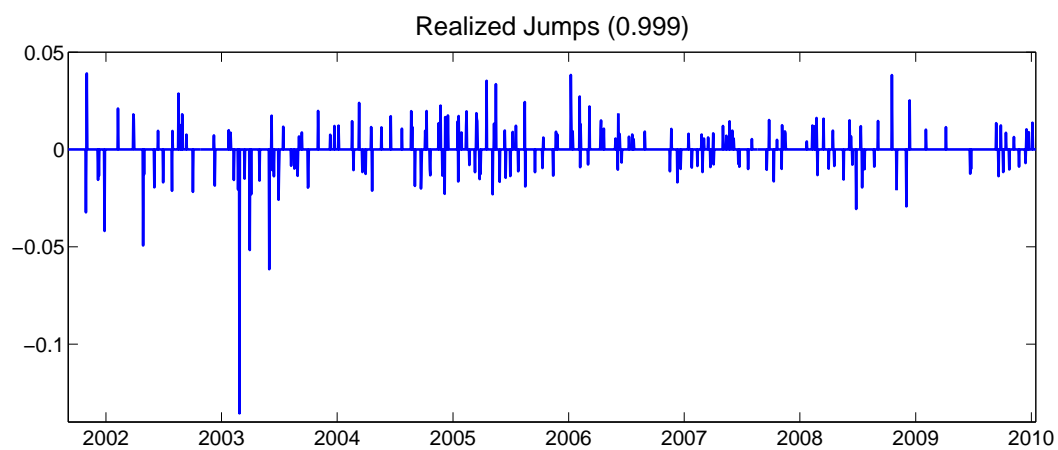
**Figure 1**  
Closing prices for the period 2001, 5th October to 2010, 15th, January.



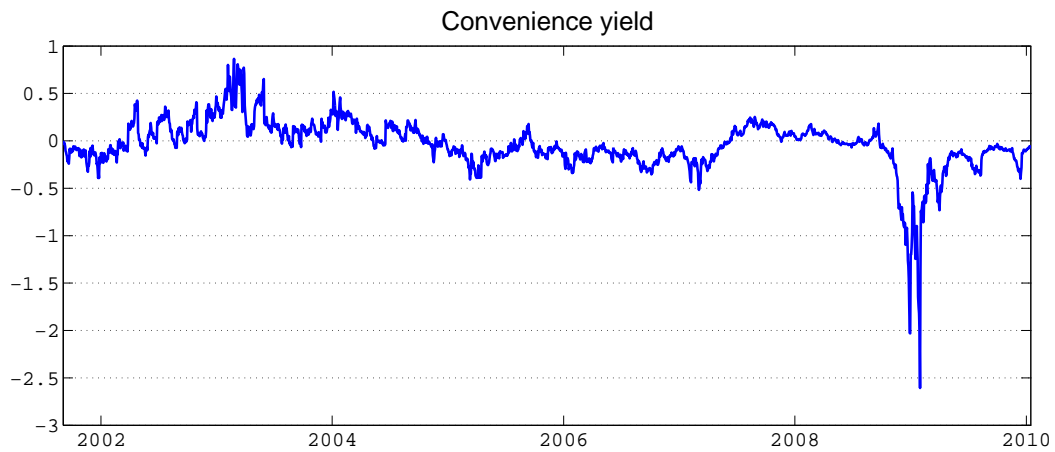
**Figure 2**  
Daily returns for the period 2001, 5th October to 2010, 15th, January.



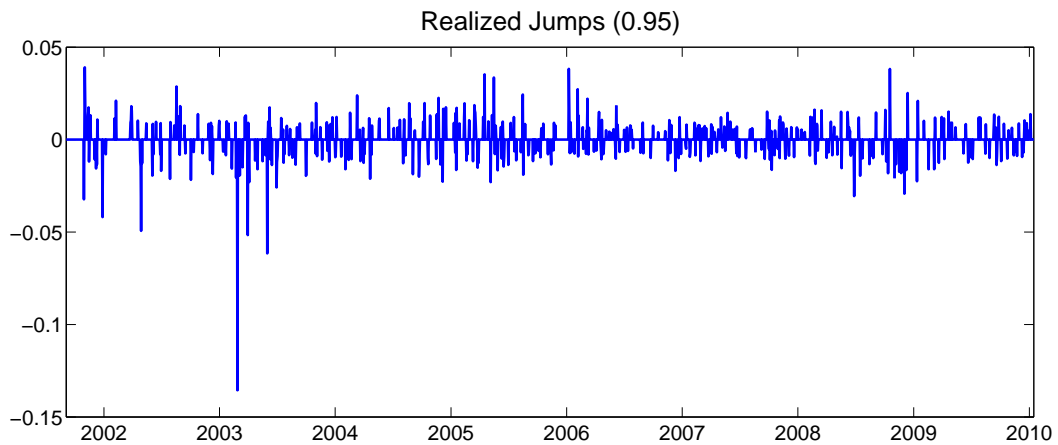
**Figure 3**  
Realized volatility (annualized) for the period 2001, 5th October to 2010, 15th, January.



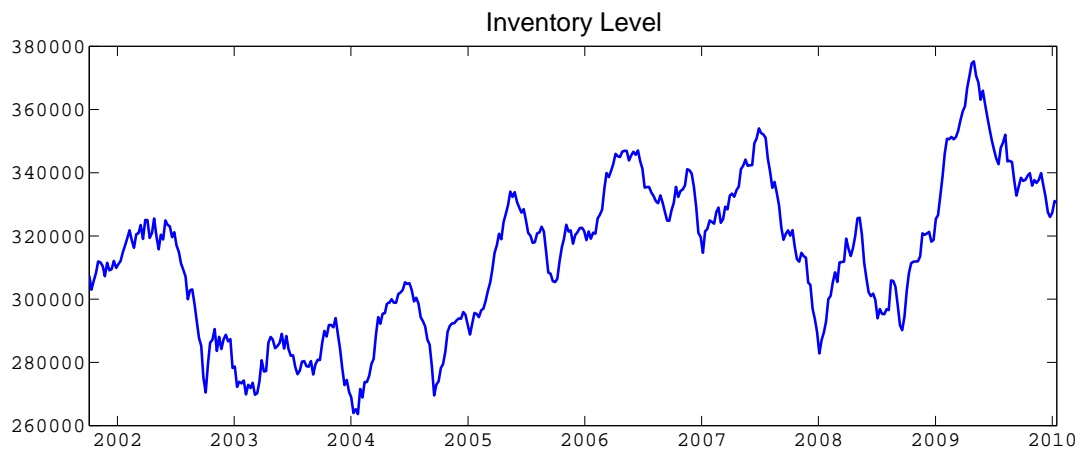
**Figure 4**  
Realized jumps for the period 2001, 5th October to 2010, 15th, January using a 0.1% detection threshold.



**Figure 5**  
Convenience yield computed using Eq.(7).



**Figure 6**  
Realized jumps for the period 2001, 5th October to 2010, 15th, January using a 5% detection threshold.



**Figure 7**  
Crude oil inventory level (weekly data) using EIA source.

## Tables

	CY	d(inventory)	5-week volatility (daily data)	8-week volatility (tick data)	Jump Mean (0.001, 250 days)	Jump Volatility (0.001, 250 days)
Mean	-0.0441	11881	1.5440	2.4254	-0.3097	1.7870
Median	-0.0381	18800	0.9900	1.6219	0.8186	1.6349
Std. dev.	0.2713	354804	1.7062	2.5923	6.1598	0.7030
Minimum	-2.5733	-998200	0.2577	0.7053	-16.7814	0.9302
Maximum	0.7711	939000	10.3555	15.3618	8.2174	3.4840
$\rho_1$	0.0963	0.0627	0.9697	0.9844	0.9898	0.9801
$\rho_2$	0.0756	0.0220	0.9288	0.9678	0.9816	0.9633
$\rho_3$	0.0636	0.0311	0.8874	0.9443	0.9738	0.9470
$\rho_{10}$	0.0499	-0.0426	0.6459	0.6908	0.8950	0.8383
Nb Obs.	382	382	382	382	382	382

**Table 1**

Descriptive statistics for the convenience yield (CY), the first difference of inventory level (d(inventory)), the 5-week volatility computed using daily data, the 8-week volatility computed using intraday data, the jump mean and the jump volatility calculated following Section 2.2. The period of interest is October 2001 to January 2010 and the frequency for observations is weekly.



Explanatory variables	1	2	3	4	5
constant	1.9749 (7.0948)	0.0857 (2.4615)	0.1084 (3.2548)	-0.0495 (-1.7299)	-0.2553 (-4.2814)
inventory level(-1)	-6.44E-06 (-6.9982)				
5-week volatility (daily data)		8.1733 (3.3582)			
8-week volatility (tick data)			6.0068 (4.5818)		
Jump Mean (0.1%, 250 days)				1.7601 (7.8090)	
Jump Volatility (0.1%, 250 days)					11.824 (4.0690)
Adjusted R2	0.3546	0.2627	0.3257	0.1573	0.0914

**Table 2**

The regressions are of the form:

$$CY_t = c + \alpha d(\text{Inventories}_{t-1}) + \beta \hat{\sigma}_{t-1} + \gamma \widehat{\text{Jumps}}_{t-1} + \varepsilon_t$$

with  $CY_t$  computed at time  $t$ ,  $c$  a constant,  $\text{Inventories}_{t-1}$  the level of inventories as released by the EIA,  $\widehat{\text{Jumps}}_{t-1}$  an estimate of a characteristic of the distribution of jumps and  $\hat{\sigma}_{t-1}$  an estimate of the volatility over the last 5 weeks using daily returns or over the last 8 weeks using realized volatility (intraday data with overnight returns). Newey-West adjusted  $t$ -statistics are reported in parenthesis.

Explanatory variables	1	2	3	4	5	6
constant	-0.0145 (-0.1658)	1.8641 (4.9517)	1.9583 (5.5199)	1.8095 (6.7588)	0.1084 (3.2783)	1.4461 4.5583
inventory level(-1)		-6.09E-06 (-4.9184)	-6.43E-06 (-5.9762)	-5.51E-06 (-6.3894)		-4.35E-06 -4.2450
5-week volatility (daily data)					4.2739 (1.0936)	
8-week volatility (tick data)				5.0315 (3.7965)	8.6822 (3.5024)	5.3160 (4.0002)
Jump Mean (0.1%, 250 days)	1.9442 (4.5261)	-0.2999 (-0.9763)				0.8378 (2.4807)
Jump Volatility (0.1%, 250 days)	-1.9899 (-0.3866)		0.6479 (0.2196)			
Adjusted R2	0.1560	0.3734	0.3706	0.5773	0.3315	0.6131

**Table 3**

The regressions are of the form:

$$CY_t = c + \alpha d(\text{Inventories}_{t-1}) + \beta \hat{\sigma}_{t-1} + \gamma \widehat{\text{Jumps}}_{t-1} + \varepsilon_t$$

with  $CY_t$  computed at time  $t$ ,  $c$  a constant,  $\text{Inventories}_{t-1}$  the level of inventories as released by the EIA,  $\widehat{\text{Jumps}}_{t-1}$  an estimate of a characteristic of the distribution of jumps and  $\hat{\sigma}_{t-1}$  an estimate of the volatility over the last 5 weeks using daily returns or over the last 8 weeks using realized volatility (intraday data with overnight returns). Newey-West adjusted  $t$ -statistics are reported in parenthesis.

Explanatory variables	1	2	3	4	5	6	7	8	9
constant	0.0857 (2.4615)	0.0466 (1.9756)	0.0616 (2.2238)	0.0761 (2.4666)	0.0904 (2.6307)	0.0976 (2.8183)	0.0999 (3.0063)	0.1034 (3.1317)	0.1084 (3.2548)
5-week volatility (daily data)	8.1733 (3.3582)								
1-week volatility (tick data)		28.7377 (3.4508)							
2-week volatility (tick data)			16.734 (3.3061)						
3-week volatility (tick data)				12.6638 (3.4062)					
4-week volatility (tick data)					10.6519 (3.5084)				
5-week volatility (tick data)						8.9634 (3.7854)			
6-week volatility (tick data)							7.5766 (4.1452)		
7-week volatility (tick data)								6.6414 (4.3815)	
8-week volatility (tick data)									6.0068 (4.5818)
Adjusted R2	0.2627	0.1813	0.2010	0.2371	0.2771	0.3007	0.3036	0.3113	0.3257

**Table 4**

The regressions are of the form:

$$CY_t = c + \beta \hat{\sigma}_{t-1} + \varepsilon_t$$

with  $CY_t$  computed at time  $t$ ,  $c$  a constant and  $\hat{\sigma}_{t-1}$  an estimate of the volatility over the last  $n$  weeks before  $t$  with  $n = 1, \dots, 8$ . Newey-West adjusted  $t$ -statistics are reported in parenthesis.

## References

- ANDERSEN, T.G., BENZONI, L., 2009. Realized volatility. In: Andersen, T.G., Davis, R.A., Kreiß, J.-P., Mikosch, Th. (Eds.) *Handbook of Financial Time Series*, Springer.
- ANDERSEN, T.G., BOLLERSLEV, T., DIEBOLD, F.X., 2007. Roughing it up: including jump components in the measurement, modeling and forecasting of return volatility. *Review of Economics and Statistics* 89, 701-720.
- ANDERSEN, T.G., DOBREV, D., SCHAUMBURG, E., 2012. Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics* 169, 75-93.
- ANDERSON, R.W., DANTHINE, J.-P., 1981. Cross hedging. *Journal of Political Economy* 89, 1182-1196.
- ASKARI, H., KRICHENE, N., 2008. Oil price dynamics (2002 – 2006). *Energy Economics* 30, 2134-2153.
- BARNDORFF-NIELSEN, O., SHEPHARD, N., 2004. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2, 1-37.
- BARNDORFF-NIELSEN, O., SHEPHARD, N., 2006. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics* 4, 1-30.
- BAUM, C., ZERILLI, P., 2013. Spikes and stochastic volatility in commodity prices: evidence from crude oil futures prices using conditional moments of integrated volatility. Unpublished manuscript.
- BORAK, S., HÄRDLE, W., TRÜCK, S., WERON, R., 2006. Convenience yields for CO2 emission allowance futures contracts. SFB 649 Discussion Paper 2006-76, Humboldt-Universität zu Berlin.
- CASASSUS, J., COLLIN-DUFRESNE, P., 2005. Stochastic convenience yield implied from commodity futures and interest rates. *Journal of Finance* 60, 2283-2331.
- CHAN, W.H., MAHEU, J.M., 2002. Conditional Jump Dynamics in Stock Market Returns. *Journal of Business and Economic Statistics* 20, 377-389.
- CHEVALLIER, J., SÉVI, B., 2012. On the volatility-volume relationship in energy futures markets using intraday data. *Energy Economics* 34, 1896-1909.
- CHIOU WEI, S.Z., ZHU, Z., 2006. Commodity convenience yield and risk premium determination: The case of the U.S. natural gas market. *Energy Economics* 28, 523-534.
- DRIESPRONG, G., JACOBSEN, B., MAAT, B., 2008. Striking oil: Another puzzle? *Journal of Financial Economics* 89, 307-327.
- DUMITRU, A.-M., URGA, G., 2012. Identifying jumps in financial assets: A comparison between nonparametric jump tests. *Journal of Business and Economic Statistics* 30, 242-255.
- EVANS, K.P., 2011. Intraday jumps and US macroeconomic news announcements. *Journal of Banking and Finance* 35, 2511-2527.
- GIBSON, R., SCHWARTZ, E.S., 1990. Stochastic convenience yield and the pricing of oil contingent claims. *Journal of Finance* 45, 959-976.
- GIOT, P., LAURENT, S., PETITJEAN, M., 2010. Trading activity, realized volatility and jumps. *Journal of Empirical Finance* 17, 168-175.
- GORTON, G.B., HAYASHI, F., ROUWENHORST, K.G., 2013. The fundamentals of commodity futures returns. *Review of Finance* 17, 35-105.
- GOSPODINOV, N., NG, S., 2013. Commodity prices, convenience yields, and inflation. *Review of Economics and Statistics* 95, 206-219.
- GRONWALD, M., 2012. A characterization of oil price behavior – Evidence from jumps models. *Energy Economics* 34, 1310-1317.
- HILLIARD, J.E., REIS, J., 1998. Valuation of commodity futures and options under stochastic convenience yields, interest rates, and jump diffusions in the spot. *Journal of Financial and Quantitative Analysis* 33, 61-86.
- HUANG, X., TAUCHEN, G., 2005. The relative price contribution of jumps to total price variance. *Journal of Financial Econometrics* 3, 456-499.
- KNITTEL, C.R., PINDYCK, R.S., 2013. The simple economics of commodity price speculation. CEEPR Working Paper 2013-006, Massachusetts Institute of Technology, Cambridge
- LEE, Y.-H., HU, H.-H., CHIOU, J.-S., 2010. Jump dynamics with structural breaks for crude oil prices. *Energy Economics* 32,

- LIU, L., PATTON, A.J., SHEPPARD, K., 2012. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. Unpublished manuscript.
- LIU, P., TANG, K., 2011. The stochastic behavior of commodity prices with heteroskedasticity in the convenience yield. *Journal of Empirical Finance* 18, 211-224.
- LIU, L., WAN, J., 2012. A study of Shanghai fuel oil futures price volatility based on high frequency data: Long-range dependence, modeling and forecasting. *Economic Modelling* 29, 2245-2253.
- MIRANTES, A.G., POBLACIÓN, J., SERNA, G., 2013. The stochastic seasonal behavior of energy commodity convenience yields. *Energy Economics* 40, 155-166.
- PIERONI, L., RICCIARELLI, M., 2008. Modelling dynamic storage function in commodity markets: Theory and evidence. *Economic Modelling* 25, 1080-1092.
- PINDYCK, R.S., 1994. Inventories and short run dynamics of commodity prices. *RAND Journal of Economics* 25, 141-159.
- PINDYCK, R.S., 2001. The dynamics of commodity spot and futures markets: A primer. *Energy Journal* 22, 1-29.
- PINDYCK, R.S., 2004. Volatility and commodity price dynamics. *Journal of Futures Markets* 24, 1029-1047.
- POLLET, J., 2003. Predicting asset returns with expected oil price changes. Working paper.
- SCHWARTZ, E.S., 1997. The stochastic behavior of commodity prices: Implications for valuation and hedging. *Journal of Finance* 52, 923-973.
- SCHWARTZ, E.S., SMITH, J.E., 2000. Short-term variations and long-term dynamics in commodity prices. *Management Science* 46, 893-911.
- SÉVI., B., 2014. Forecasting the volatility of crude oil futures using intraday data. *European Journal of Operational Research* 235, 643-659.
- SYMEONIDIS, L., PROKOPCZUK, M., BROOKS, C., LAZAR, E., 2012. Futures basis, inventory and commodity price volatility: An empirical analysis. *Economic Modelling* 29, 2651-2663.
- TAUCHEN, G., ZHOU, H., 2010. Realized jumps on financial markets and predicting credit spreads. *Journal of Econometrics* 160, 102-118.
- THEODOSIOU, M., ŽIKEŠ, P., 2011. A comprehensive comparison of alternative tests for jumps in asset prices. Working Paper, Central Bank of Cyprus, no. 2.
- TSENG, T.-C, CHUNG, H., HUANG, C.-S, 2009. Modeling jump and continuous components in the volatility of oil futures. *Studies in Nonlinear Dynamics & Econometrics* 13, Issue 3, Article 5.
- WANG, T., WU, J., YANG, J., 2008. Realized volatility and correlation in energy futures markets. *Journal of Futures Markets* 28, 993-1011.
- WRIGHT, J.H., ZHOU, H., 2009. Bond risk premia and realized jump risk. *Journal of Banking and Finance* 33, 2333-2345.
- ZHANG, L., MYKLAND, P.A., AÏT-SAHALIA, Y., 2005. A tale of two time scales: determining integrated volatility with noisy high frequency data. *Journal of the American Statistical Association* 100, 1394-1411.
- ZHANG, B.Y., ZHOU, H., ZHU, H., 2009. Explaining credit default swap spreads with the equity volatility and jump risks of individual firms. *Review of Financial Studies* 22, 5099-5131.